I discuss various thoughts, old and new, about the cosmological constant (or dark energy) paradox. In particular, I suggest the possibility that the cosmological "constant" may decay as $\Lambda \sim \alpha^2 m_N^3/\tau$, where τ is the age of the universe.

I. THE PARADOX

Physics thrives on paradoxes. The observation of dark energy, [1] [2] which may or may not be a cosmological constant, has spread despondency and elation around the theoretical physics community, depending on the personality of the theorist. The cosmological constant, [3] [4] that phantom of Einstein's mind, turned out to be real after all: tiny in the fundamental units of the particle theorists, substantial in the natural units of the cosmologists, but most definitely not the mathematicians's zero. The best and the brightest in the theoretical community missed a golden opportunity much as the opportunity presented to young Schwinger to calculate the anomalous magnetic moment of the electron. A brilliant young theorist of our era could have been just as stylish as Schwinger and calculated, before the observers did their wonderful work, the cosmological constant to say 3 significant figures, which then turned out to be in almost exact agreement with the observed value. Alas, that would-be could-be triumph of string theory did not occur. By the way, before Schwinger came along, some theorists, equally authoritative as some of our esteemed colleagues today, also insisted in print that the anomalous magnetic moment is identically zero. [5]

As is well known, the paradox can be easily described. The natural value of Λ in particle physics is expected by dimensional analysis to be $\mu^4 = \mu/(\mu^{-1})^3$ for some relevant mass scale μ where the second form of writing μ^4 reminds us that Λ is a mass or energy density. Whether one associates μ with grand unification, electroweak symmetry breaking, or the quark confinement transition and consequently has a value of order 10^{19} Gev, 10^2 Gev, and 1 Gev respectively is immaterial. The natural value $\Lambda \sim \mu^4 = \mu/(\mu^{-1})^3$ is outrageous even if we take the smallest value for μ. We don't even have to put in actual numbers to see that there is a humongous discrepancy between theoretical expectation and observational reality. We know the universe is not permeated with a mass density of order of 1 Gev on every cube of size 1 $(Gev)^{-1}$.

The cosmological constant paradox is basically an enormous mismatch between the units natural to particle physics and natural to cosmology. Measured in units of Gev^4 the cosmological constant is so incredibly tiny that particle physicists have traditionally assumed that it must be mathematically zero, and have looked in vain for a plausible mechanism to drive it to zero.

But Nature has a big surprise for us. While theorists racked their brains trying to come up with a convincing argument that $\Lambda = 0$, observational cosmologists [1] [2] steadily refined their measurements and changed their upper bound to an approximate equality [6]

$$\Lambda \sim (10^{-3}ev)^4 !!! \tag{1}$$

Thus, theorists once expected the cosmological constant to be "naturally" humongous, then insisted that it must be mathematically zero, in spite of hints [7] that it may not be observationally zero. Now observers have established that it is in fact tiny but not zero.

As is also well known, we should, strictly speaking, refer to the observation of the cosmological constant as the observation of a hiherto unknown dark energy since we do not know the equation of state associated with the observed energy density $(10^{-3}ev)^4$. In this article, we use the terms "cosmological constant" and "dark energy" interchangeably.

To make things worse, the energy density $(10^{-3}ev)^4$ happens to be the same order of magnitude as the present matter density of the universe ρ_M. This is sometimes referred as the cosmic coincidence problem. The cosmological constant Λ is, within our traditional understanding, a parameter in the Lagrangian. On the other hand, since most of the mass density of the universe resides in the rest mass of baryons, $\rho_M(t)$ decreases like $(1/R(t))^3$ as the universe expands where $R(t)$ denotes the scale size of the universe. In the far past, ρ_M was much larger than Λ, and in the far future, it will be much smaller. It just so happens that in this particular epoch of the universe, when we are around, that $\rho_M \sim \Lambda$. Or to be less anthropocentric, the epoch when $\rho_M \sim \Lambda$ happens to be when galaxy formation has been largely completed. In their desperation, some theorists have even been driven to invoke anthropic selection [8] [9] [10].

In writing this article for these proceedings, I went back and re-read a talk [4] on the cosmological constant problem I gave just over 20 years ago. To my consternation, and utter humility, I find that many of the thoughts now current in the theoretical community, including the proton decay analogy, were already discussed or at least alluded to in that talk. Incidentally, the talk was given at a conference "in Honor of P. A. M. Dirac in his Eightieth Year". [11] Unhappily that was the last occasion I saw Dirac. He died on October 20, 1984 before the proceeding could be

published. As another historical note, J. Schwarz gave a talk on string theory right after my talk, and since this was before the so-called string revolution his talk was rather sparsely attended. I remember quite clearly having lunch with him after his talk and his admonition to me that I would never understand the cosmological constant without string theory. Be that as it may, I am glad that John triumphed over his adversaries, namely all those who left the lecture hall as he started talking about this at-the-time far-out thing called string theory.

In this article, I would like to highlight two suggestions I made at that conference. One is the suggestion that the cosmological "constant" depends on the age of the universe. The other is a potentially fruitful analogy with the story of proton decay. The outline of this article is as follows: after discussing these two suggestions I list some of the other points I made in that talk [4], some of which may still prove to be relevant.

II. A COSMOLOGICAL CONSTANT DECAYING LINEARLY IN TIME

In that talk, since the occasion was in honor of Dirac, I was inspired by Dirac's large number hypothesis and ventured to propose for the cosmological constant the formula

$$\Lambda \sim \frac{m_N^3}{\alpha^2 \tau} \tag{2}$$

where τ denotes the age of the universe. We obtain an oxymoronic "cosmological constant" that decays with time. Recall that Dirac, faced with a retreat from the proposition that large dimensionless numbers should not appear in physics, suggested that the two large numbers $\alpha/(Gm_N^2)$ and $m_N \tau$ should be equal to each other. I obtained (2) in the same spirit, starting with the large number obtained by dividing the expected value of the cosmological constant by the supposed upper bound considered reasonable at that time. At present, the proposed formula would work better numerically if I moved the fine structure constant squared from the denominator to the numerator:

$$\Lambda \sim \frac{\alpha^2 m_N^3}{\tau} \tag{3}$$

If I take the age of the universe τ to be 15×10^{10} years $= 4.7 \times 10^{17}$ seconds which corresponds to an energy of 1.4×10^{-33} electron volts, the proposed formula (2) gives

$$\Lambda \sim (3 \times 10^{-3} ev)^4 \tag{4}$$

Of course, numerically, it is much less sensitive to quote the mass scale of the physics responsible for the cosmological constant than to quote the cosmological constant itself. As long as we are playing around we might note the alternative formulas

$$\Lambda \sim \frac{m_e^3}{\tau} \tag{5}$$

which gives $\Lambda \sim (10^{-4} ev)^4$ and

$$\Lambda \sim \frac{(\alpha_s m_N)^3}{\tau} \tag{6}$$

with α_s the strong interaction coupling constant. Needless to say, this is not serious physics and should be taken as exploratory at best. I am merely suggesting that Dirac's large number hypothesis appears to lead to a dark energy density decaying linearly with time.

In the matter dominated era $(\frac{\dot{R}}{R})^2 \propto \rho_M$ and so the matter density ρ_M scales with time like $\rho_M \propto 1/t^2$. Thus, a dark energy density decaying linearly with time would not affect the rather impressive results on primordial nucleosynthesis.

III. THE PROTON DECAY ANALOGY

The story of the cosmological constant can be summarized as follows: it is a measurable physical quantity once thought to have the value $\lesssim \infty$, then thought to be 0, and finally measured to have the value $\neq 0$ but $\lll \infty$.

I asked myself, "Has there ever been a physical quantity that has gone through such a dramatic double 'reversal of fortune'?"

I came up with the proton decay rate. Recently, I wrote a speculative paper [12] pointing out this historical precedent and suggesting that it may shed some light on the resolution of the cosmological constant paradox. Remarkably, when I wrote that paper I totally forgot that I had already published it in the proceedings of my 1983 talk [4] and referred only to my field theory textbook. [13]

Let us go through the story of proton decay. To make my point I will take some liberty with history. Suppose that in 1953 some theorists were to calculate the rate Γ for protons to decay in the natural mode $p \to e^+ + \pi^0$. The interaction of the pion with the proton and the neutron was known to be described by a term like $g\pi\overline{n}p$ in the Lagrangian with g a dimensionless coupling of order 1. These theorists would naturally construct a Lagrangian out of the available fields, namely the proton field p, the electron field e, and the pion field π, and thus write down something like $f\pi\overline{e}p$ with some constant f. Note that $\pi\overline{e}p$ has mass dimension 4 and hence f is dimensionless just like g. Since $\pi\overline{e}p$ violates isospin invariance, the theorists would expect f to be suppressed relative to g by some measure of isospin breaking, say the fine structure constant α. The natural value for Γ would then come out to be many many orders of magnitude larger than the experimental upper bound on Γ. The theorists would then set $\Gamma = 0$ and cast about for an explanation. After an enormous struggle, the theorists were unable to come up with a compelling explanation and this failure became known as the proton decay rate paradox.

Eventually, someone with great authority and prestige in the community, namely Wigner, decreed the law of baryon number conservation. Surely, even in the unthinkably primitive days of 1953 this would have been recognized as a pronouncement and not as an explanation. (The pronouncement could be dressed up formally by imposing a $U(1)$ transformation under which $p \to e^{i\theta}p$ while e and π do not change and requiring that the Lagrangian remains invariant.) But there would have been no deep understanding of this astonishing discrepancy between theoretical expectation and experimental upper bound.

Indeed, imagine an alternative history in which, while other important particle physics experiments were being performed in 1957, some intrepid experimentalist, ignoring conventional theoretical wisdom, actually went out and measured the proton decay rate to be some tiny but non-zero value. The proton decay rate paradox would have deepened, much as how the cosmological constant paradox deepened with the discovery of a tiny Λ.

Let us now review how the proton decay rate paradox was resolved historically. The first remark is that the eventual explanation did not emerge within the orthodox theory fashionable in 1957, nor did it come from an understanding of some kind of mechanism causing protons to decay, but rather it came totally from "left field", from a study of baryon spectroscopy, which led to the notion of quarks. The correct degrees of freedom are not given by the proton and pion fields p and π, but by the quark fields q. The effective Lagrangian $\mathcal{L}$ is to be constructed out of quark q and lepton l fields and must satisfy the symmetries that we know. Three quarks disappear, so we write down schematically qqq, but three spinors do not a Lorentz scalar make. We have to include a lepton field and write $qqql$. Since four fermion fields are involved, these terms have mass dimension 6 and so in $\mathcal{L}$ they have to appear as $\frac{1}{M^2}qqql$ with some mass M, corresponding to the mass scale of the physics responsible for proton decay. Thus, the probability of proton decay is proportional to $(\frac{1}{M^2})^2 = \frac{1}{M^4}$. By dimensional reasoning, we obtain the proton decay rate $\Gamma \sim (\frac{m_p}{M})^4 m_p$. The absurdly small value of Γ is then naturally explained by the fourth power of the small number $(\frac{m_p}{M})$ for M big enough. No mystery left!

Note that in principle all of this could be done as soon as Gell-Mann introduced the notion of quarks in 1964, long before anybody even dreamed of a grand unified theory with proton decay.

As long we are discussing revisionist, but possible, history, we can imagine some brilliant theorist in another civilization far far away puzzling over the proton decay rate paradox eventually realizing that the key to explaining an absurdly small number is to promote the dimension of the effective Lagrangian merely from 4 to 6. In hindsight, we can say that the extremely long lifetime of the proton could have pointed to the existence of quarks.

I would like to raise the question whether the cosmological constant paradox might not be solved in the same way. Perhaps the gravitational field $g_{\mu\nu}$ is the analog of the proton and pion field p and π. The high energy and more fundamental degrees of freedom in the gravitational field may not be the metric $g_{\mu\nu}$, but some mysterious analog of the quark field q. This may emerge as a construct in string or M theory, or it could be something else completely. In the history of the proton decay paradox as recounted by me, there is an additional twist, namely that the degree of freedom q is confined and not physical. Before the advent of quantum chromodynamics, theorists could only write $p \sim qqq$, without any clear idea about what the symbol $\sim$ might mean. We are in a similar position here: the metric $g_{\mu\nu}$ might be a composite object, but I certainly do not know what it is a composite of, and the objects of which $g_{\mu\nu}$ is a composite may also be as observable or as unobservable as the quarks. Just as we did not know what the symbol $\sim$ might mean in $p \sim qqq$, we may have no idea of even what the word "composite" might mean in the case of $g_{\mu\nu}$.

The cosmological term $\Lambda\sqrt{g}$ in the Lagrangian has mass dimension 0 and we somehow have to promote 0 to a higher number. One difficulty with this view is of course how we could possibly promote the dimension of the cosmological term without at the same time changing the mass dimension of the Einstein-Hilbert term $\frac{1}{G}\sqrt{g}R$. Our historical analogy may again be helpful: the 1953 view that the pion nucleon coupling term has dimension 4 turns out to be

3

correct. While the dimension 4 term $\pi \overline{e} p$ was replaced by the dimension 6 term $qqql$ the dimension 4 term $\pi \overline{n} p$ was replaced by dimension 4 terms of the form $\overline{q} A q$ with A a gluon potential. The dimension of one of the two terms gets promoted while the dimension of the other term remains the same. So it is entirely conceivable to me that the cosmological constant term could end up with a higher dimension while the Einstein-Hilbert term either remains dimension 4 or is replaced by dimension 4 terms. Thus, suppose the cosmological constant term actually has dimension $p > 0$ so that it is given in the Lagrangian by a term of the form $\frac{1}{M^{p-4}}\mathcal{O}$ with M some mass scale characteristic of the deeper structure of the graviton, perhaps the same as the Planck mass, perhaps not. The observed cosmological constant would then be given by $\Lambda \sim \frac{1}{M^{p-4}} < \mathcal{O} > = (\frac{m}{M})^p M^4$, where the expectation value of the operator $\mathcal{O}$ in the physical universe $< \mathcal{O} > = m^p$ is set by physics at some low energy scale m. With m small enough, and or p big enough, we could easily get the suppression factor we want.

IV. NATURE OF GRAVITY

One difficulty with this analogy may be that it relies implicitly on our understanding of quantum field theory as low energy effective field theories. Banks, Fishler, and others have emphasized to me that this customary understanding may not apply to gravity. Our formulation of field theory relies on that marvellous invention known as the Fourier transform e^{ipx} according to which low momentum and energy corresponds to long distances, so that microscopic details do not enter the effective theory.

However, while this appears to govern all particles it does not appear to work for black holes. According to the celebrated relation $R = GM$ the more massive a black hole the larger it is. The size of a black hole is not governed by its Compton wavelength! Evidently, this strange result follows from the fact that gravity, in contrast to the other fundamental interactions, has a dimensionful coupling constant. The relation $R = GM$ follows from the dimension of G and the supposition that black holes should not exist as we turn off gravity (supplemented by a certain notion of simplicity.) We could imagine (although we do not know how to actually do it) modifying Einstein's theory so that this relation holds for astronomical black holes but not for microscopic black holes that we should include in our effective field theories. Perhaps G depends on M in such a way that we could recover the deBroglie-Compton relation that $R \sim M^{-1}$. Or perhaps in the standard form of the Schwarzschild solution g_{00} is modified to $(1 - \frac{GM}{f(r)})$ so that $f(r) \to r$ for large r but behaves differently for small r, so that the dependence of R on M is changed. These of course are just "empty" remarks at this point.

Another point Banks [14] has impressed upon me is a possible connection between gravity and quantum measurement theory. In the usual formulation, a classical measuring apparatus, classical because it could be made very large and massive, is needed to study a quantum system. All very well without gravity! But with gravity, the massive classical apparatus would exert an unacceptable gravitational effect on the system we are trying to measure. To minimize this perturbation we are obliged to move the classical apparatus to infinity, which according to Banks gives a whiff of 't Hooft's holographic principle. I find this very disturbing, to say the least.

While invoking the proton decay analogy I do not necessarily insist that the graviton is composite, although that is certainly a possibility. The graviton could be part of a larger structure. [15] Inspired by the almost exact correspondence between Einstein's post-Newtonian equations of gravity and Maxwell's equations of motion I once proposed the gravitipole in analogy with Dirac's magnetic monopole. After Dirac there was considerable debate on how a field theory of magnetic monopoles may be formulated. Eventually, 't Hooft and Polyakov showed that the magnetic monopole exists as an extended solution in certain non-abelian gauge theories. Most theorists now believe that electromagnetism is merely a piece of a grand unified theory and that magnetic monopoles exist. Might it not turn out that Einstein's theory is but a piece of a bigger theory and that gravitipoles exist? In grand unified theory the electromagnetic field is a component of a multiplet. Could it be that some generalized gravitational field $\mathcal{G}_{\mu\nu}^{\alpha}$ also somehow carries an internal index α and that the gravitational field we observe is just a component of a multiplet? Throwing caution to the wind, I also asked in [15] if the gravitipole and the graviton might not form a representation under some dual group just as the magnetic monopole and the photon form a triplet under the dual group of Montonen and Olive [16].

In this connection, I have always been quite taken by 't Hooft's elegant expression for Maxwell's electromagnetic field

$$\mathcal{F}_{\mu\nu} = \frac{F_{\mu\nu}^a \varphi^a}{|\varphi|} - \frac{\varepsilon^{abc}\varphi^a (D_\mu \varphi)^b (D_\nu \varphi)^c}{|\varphi|^3}$$

in terms of the Yang-Mills field. I wonder from time to time if there might not be an analogous expression, giving Einstein's gravitational field $g_{\mu\nu}$ in terms of $\mathcal{G}_{\mu\nu}^{\alpha}$.

4

V. MISCELLANEOUS THOUGHTS

There is of course no lack of speculative thoughts [17] about the cosmological constant in the literature. Let me take this opportunity to list some of the other points I mentioned in my talk [4].

(1) The fact that gravity is not renormalizable is not really relevant as long as we wish to study only gravitational physics at low energy and long distances. We have an effective low energy theory consisting of an expansion in powers of $(1/M_{Pl})^2$ and with radiative corrections cut off at some unknown mass scale M_C. The cosmological constant and the Einstein-Hilbert action correspond to the first two terms.

(2) The graviton knows about anything carrying energy and momentum, including an apparently innocuous constant shift in the Lagrangian density $\mathcal{L} \to \mathcal{L} - \Lambda$. One possible way of nullifying the physical consequence of the shift $\mathcal{L} \to \mathcal{L} - \Lambda$ is to postulate that $g = \det g_{\mu\nu}$ is not a dynamical variable. This was proposed [18] [19] as an explanation of why Λ is mathematically zero. But with Λ now known to be tiny but non-zero this avenue seems to me less promising. Historically, it is amusing to note that Einstein did his first calculation [20] imposing the gauge choice $g = 1$. This approach has been explored by S. Weinberg [3] and others.

(3) Besides the proton decay analogy mentioned above, I also pointed out 2 other possibly relevant analogies. I mention one of them here. In the days when the strong interaction was described by the Yukawa theory, it was mysterious why the strong interaction is parity and charge conjugation invariant and conserves strangeness. One can perfectly well introduce terms such as $a\overline{\Sigma}p$ and $m_N\overline{p}(1 + b\gamma_5)p$ (here Σ denotes the sigma hyperon) and so forth into the Lagrangian. The "natural" expectation would be that the dimensionless parameters a and b are of order unity, while experimentally they are known to be very small. Even if we had a reason to exclude terms like these they would be induced by the weak interaction. At that time physicists were at a loss to explain why these terms are not there. The solution did not come until the formulation of quantum chromodynamics. Gauge invariance allows the offending terms to be simply transformed away. As forcefully emphasized by Weinberg, this represents one of the virtues of quantum chromodynamics.

This example may be telling us that perhaps Einstein's theory is to the true theory of gravity as Yukawa's theory is to quantum chromodynamics.

(4) I discussed the possibility of promoting Λ into a dynamical field $\Lambda(x)$, which I mentioned is tantamount to introducing a scalar field $\varphi(x)$. Somehow $\varphi(x)$ would move dynamically to counter any "bare" Λ. To do this, the scalar field must know about Λ, and thus must be coupled to the curvature in various ways. I believe that some versions of this kind of approach is captured in the contemporary literature in the notion of "tracker fields".

(5) The cosmological constant is a property of the particular ground state of the Universe at a particular instant. For example, in induced gravity [21] [22] Λ is given by $< 0|T|0 >$ and thus depends on the ground state $|0 >$ (here T denotes the trace of the stress energy tensor of course.) However, the cosmological evolution $|0 >$ depends on Λ, schematically, $\frac{\partial}{\partial t}|0 >= F[\Lambda]$ for some functional F, so that $\frac{\partial \Lambda}{\partial t}$ is a functional of Λ. It is thus conceivable that the cosmological "constant" might relax to zero or near zero in the present universe. This approach was pursued by Adler, Myhrvold, Mottola, and others without reaching a definitive conclusion.

(6) There is always the possibility [23] that the quantum field theory formulation of the "vacuum" is incorrect or inconvenient. In modern times, this is perhaps represented by the thoughts behind the holographic principle as outlined above by Banks. But I am rather disinclined not to trust quantum field theory; at the very least, quantum field theory has already survived two near-death experiences.

Perhaps we do not know as much about the graviton as we think we do. Possibly one concrete sign of this is the amazing discovery by Bern [24] and others that the Einstein-Hilbert action may factorize order by order in the gravitational field into what one might think of as a product of two Yang-Mills actions.

VI. REFERENCES

[1] A. G. Riess et al, Astron. J. 116, 1009 (1998)
[2] S. Perlmutter et al, Astrophys. J. 517, 565 (1999)
[3] S. Weinberg, Rev. Mod. Phys. 61, 1 (1989)

[4] A. Zee, "Remarks on the Cosmological Constant Paradox," in "High Energy Physics" in Honor of P. A. M. Dirac in his Eightieth Year, edited by S. L. Mintz and A. Perlmutter, Proceedings of the 20th Orbis Scientiae (1983), p. 211, Plenum Press, New York 1985.

[5] D. Kosower, private communication.

[6] The value 10^{-3} ev does not correspond to any known mass scale in particle physics. The differences in mass squared of the 3 neutrinos have been measured experimentally to be $|m_3^2 - m_2^2| \sim 3 \times 10^{-3} ev^2$ and $|m_2^2 - m_1^2| \sim 7 \times 10^{-5} ev^2$ and so it is conceivable that the lightest neutrino has mass $\sim 10^{-3}$ ev and that this may have something to do with the cosmological constant. In a recent paper, using a particular Ansatz X. G. He and A. Zee (hep-ph/0302201) determined neutrino masses to lie in the range 10^{-2} to 10^{-4} ev.

[7] P. J. E. Peebles and B. Ratra, Ap. J. 325, L17, 1988, and references therein.

[8] T. Banks, Nucl. Phys. B249, 332 (1985)

[9] S. Weinberg, Phys. Rev. Lett. 59, 2607 (1987), Phys. Rev. D61, 103505 (2000)

[10] A. Vilenkin, Phys. Rev. Lett. 74, 846 (1995)

[11] "High Energy Physics" in Honor of P. A. M. Dirac in his Eightieth Year, edited by S. L. Mintz and A. Perlmutter, Proceedings of the 20th Orbis Scientiae (1983), Plenum Press, New York 1985.

[12] A. Zee, "Dark Energy and the Nature of the Graviton", hep-th/0309032.

[13] See page 440 in A. Zee, "Quantum Field Theory in a Nutshell", Princeton University Press, April 2003.

[14] T. Banks, private communication.

[15] A. Zee, Phys. Rev. Lett. 55, 2379 (1985)

[16] C. Montonen and D.Olive, Phys. Lett. 72B, 117 (1977)

[17] See for example, R. Sundrum, hep-th/0306106, and references therein.

[18] F. Wilczek and A. Zee, unpublished, see discussion in [4]. This possibility was analyzed further in section VI of [3].

[19] F. Wilczek, Phys.Rev.Lett. 80 (1998) 4851.

[20] A. Einstein, Ann. der Physik 49 (1916). Note specifically the discussion around eqs. (9, 44,47, & 47a). In particular, $R_{\mu\nu}$ simplifies tremendously.

[21] S. Adler, Rev. Mod. Phys. 54, 729 (1982).

[22] A. Zee, Phys. Rev. D23, 858 (1981), Phys. Rev. Lett. 48, 295 (1982), " Gravity as a Dynamical Consequence of the Strong, Weak and Electromagnetic Interactions" in "Unification of the Fundamental Particle Interactions", Proceedings of the Erice Conference 1981, (Plenum Press), Ann. of Phys. 151, 431 (1983)

[23] R. P. Feynman, private communication.

[24] Z. Bern, gr-qc/0206071, and references cited therein.

Acknowledgments